Information Systems Division
CO-OPS Technical Plan for Data Exchange and Dissemination

Mohamed Chaouchi

Silver Spring, Maryland
July 2005

noaa National Oceanic and Atmospheric Administration

U.S. DEPARTMENT OF COMMERCE
National Ocean Service
Center for operational Oceanographic Products and services / Information Systems Division

TABLE OF CONTENTS

1. Introduction

The Center for Operational Oceanographic Products and Services (CO-OPS) is supporting NOAA's effort in the implementation of the Integrated Ocean Observing System (IOOS). The Ocean.US (http://ocean.us/) Data Management and Communications (DMAC) Plan gives general guidance for implementation, which CO-OPS will use to implement a set of applications to address its specific needs and requirements in fulfillment of the IOOS vision. This technical plan describes the goals and their associated objectives and milestones. In addition, this plan recommends technologies and strategies to help reach these milestones.

By complying with IOOS, CO-OPS will advance the core of its broader mission: to serve the information needs of researchers, scientists, technical employees and the public. It will improve the way CO-OPS stores, describes, transports and provides data to the users, ultimately increasing the potential for discovery, interoperability and use of this data.

2. Implementation Overview

CO-OPS is aiming to build the capability to serve model NetCDF (Network Common Data Format) files, develop Web Services and create basic IOOS-compliant metadata. The Web Services will allow easier access and transport of the metadata and data files. By having IOOS-compliant metadata, such as information about each station, it will be easier to search for a particular data item.

Initially, the CO-OPS' IOOS team will install an OPeNDAP server to serve model NetCDF files and subsequently, other data formats. It will also develop Web applications to serve as interfaces to the OPeNDAP server, Web Services, and to the metadata so that users can more easily browse, search and view CO-OPS datasets.

The CO-OPS' IOOS team will implement on-demand Web Services for a limited set of high-priority applications that will enable specific users to efficiently retrieve time-critical data. Meanwhile, the team will gather requirements to set up an entire suite of Web Services to make available all CO-OPS products.

Finally, the CO-OPS' IOOS team will formalize the metadata --data about data-- by identifying areas with insufficient information, choosing metadata standards for the observational parameters and making these standards available for discovery. These standards will be in an XML (Extensible Markup Language) format, which displays on a standard browser, allowing end-users to easily browse and use the data.

This technical plan proposes taking incremental steps toward becoming IOOS compliant by implementing best industry practices and deploying powerful and proven technologies that can be integrated easily into any framework.

Note: some of the tasks described in this document were started before and/or in parallel with the writing of this plan as a "proof of concept" to learn more about the technology required and the implementation and administration of the system applications.

3. Goals and Objectives

CO-OPS' overarching goals are to improve services to its users by expanding the methods by which data is (1) served, (2) described and (3) represented. These goals can be achieved through consistent data transport, increased interoperability, and standardization of metadata.

- **Objective #1: Consistent Data Transport**

Data Transport is a key element in the management and delivery of data. Consistent data transport ensures security, reliability and accessibility of the data regardless of the multiplicity of platforms, protocols, networks, and data formats between the data provider and the user. A primary objective of CO-OPS is to be able to transport data in a consistent and standard way through well-established protocols following the recommendations of the DMAC community.

- **Objective #2: Increase Interoperability**

Another objective is to enable application-to-application communication. Interoperability enables information systems to bypass communication protocols and data compatibility issues, allowing different servers and their applications to recognize and talk to one another without human intervention.

- **Objective #3: Standardized Metadata**

Metadata is integral to defining, presenting and displaying data. For all CO-OPS data, having well-defined metadata and a standard format will facilitate the exchange of data across agencies and increase the understandability of data. This will allow end-users to more easily view, query, and extract information of interest.

4. Major Milestones

Milestone #1: Setting up OPeNDAP Server

To enhance the way CO-OPS serves data and to increase data accessibility, the CO-OPS' IOOS team will install an OPeNDAP server to transport, serve and query NetCDF models. After installing the OPeNDAP server, end-users of the system will have easier access to the data and they will have the capability to sub-sample data and its associated syntactic metadata. This will satisfy an urgent need to make CO-OPS NetCDF models available to the emergency management teams, namely HAZMAT and the U.S. Coast Guard.

This milestone consists of four major components: acquiring a web server, choosing an OPeNDAP server, and providing its content and its interface.

- **Acquiring the Web Server –OPeNDAP Host**

The CO-OPS' IOOS team recommends Apache as the web server to host OPeNDAP. Apache is one of the most stable web servers. It is an open source project with the support of a large community of developers and users. Apache also works well with Tomcat, another Apache project and a key component that, *virtually*, all applications to be implemented will need. Tomcat is primarily a container and a servlet engine that translates JAVA code to HTML pages so browsers can display its contents, and provides other important functionalities for web development. Tomcat also runs as a stand-alone web server.

- **Choosing the OPeNDAP Server -- THREDDS**

OPeNDAP comes in a variety of servers, each specializing in serving one and sometimes several data formats. A key step in deciding which server to select involves examining the various database storage formats used by CO-OPS. Another step will be to decide which CO-OPS data sets to serve initially, and then choosing an OPeNDAP server that best serves this data.

Based on careful analysis of in-house priorities concerning the type of the data CO-OPS needs to serve, the CO-OPS' IOOS team recommends that THREDDS (Thematic Real-Time Environmental Distributed Data Services) be installed as the OPeNDAP server since it has the ability to serve model NetCDF files.

THREDDS will also allow CO-OPS to create and serve logical views of datasets. These views are aggregations and catalogs composed of several NetCDF files grouped together. In addition, THREDDS is a JAVA application that can run on Tomcat as a stand-alone server or as part of the Apache web server. Another advantage is that THREDDS has the capability to serve data generated by other OPeNDAP servers in addition to external data residing on different server(s). In the event that another OPeNDAP server needs to be installed in order to transport other types of data formats (such as OPeNDAP/DODS Java 1.1.5, which serves data that is stored in a relational database system), that data can also be served through THREDDS.

Moreover, the next release of the THREDDS server, according the UNIDATA support group, will be able to automatically detect the newly updated models and serve them in real time. This new functionality should be available by late July. This functionality is crucial since CO-OPS model data are updated hourly, and if the OPeNDAP server can detect the updated model files automatically, this will greatly simplify development time. The ability to detect new files on the fly will be a critical priority for emergency response teams such as NOAA HAZMAT and the U.S. Coast Guard. In the current version of THREDDS, it is necessary to supply a new configuration of the newly updated dataset, and then restart the server in order for the server to see the updated files.

- **OPeNDAP Content**

CO-OPS will need to specify the datasets served through THREDDS, how often they will be updated, and the information to display for the model NetCDF files. Later, CO-OPS will need to identify other datasets that will be served through OPeNDAP.

There are four files associated with each set of data, and most of them will be automatically generated. The information file, which is an HTML page, can be modified to include any metadata or other information to be displayed to the end-users. The following is brief description of these files:

(1) Dataset Descriptor Structure (DDS)

This file contains a description of the structure of the data and uses syntax similar to a higher programming language such as C. In the case of a relational database, this file will describe the table. For NetCDF files, the DDS file is automatically generated. In the case of data residing in a relational database, it will need to be supplied.

(2) Data Attribute Structure (DAS)

This file contains information about the data, such as the name of the variable, its unit and its type. In the case of a relational database, this file will describe the attributes of the table. For NetCDF files, the DDS file is automatically generated. In the case of data residing in a relational database, it will need to be supplied.

(3) Information

This HTML page has detailed information about the data in a more readable format. Unlike the two previously mentioned files, the information file is optional; however, the more information that is supplied, the better the service will be to the users.

(4) HTML Data Request Form

The OPeNDAP server automatically generates this HTML form based on the input configuration file. It serves as an interface to describe the data set. It provides variable selection boxes, which allow users to enter constraints such as applying some Boolean operation on different variables of the dataset, and to build a URL of choice that will display the data they request.

- **OPeNDAP Interface**

CO-OPS needs to create a web application that will serve as an interface to the OPeNDAP server. This application will give the users the ability to choose the dataset of interest and perform any actions on the data. This web-based application will present CO-OPS services in a user-friendly format.

The recommended platform to develop this web-based application is the Java 2 platform Enterprise Edition (J2EE) technology using Java Server Pages (JSP), Servlets, JavaBeans, and Java Database Connectivity (JDBC).

A simpler architecture may suffice to build an interface, but with this design, future expansion with more complex requirements will make the transition easier.

Milestone #2: Implementing Web Services

Web Services provide real-time, remote and automated access to CO-OPS data, which makes it easier for partners and end-users to use CO-OPS services. This interoperability will improve business processes by reducing the cost and time to connect to other web applications. It will also lead to more exposure of CO-OPS services by making them easier to find during data discovery.

Web Services are becoming the standard solutions in the industry for business-to-business applications since they allow for automated access to data. In addition, they are usually described through a set of APIs, which serve as documentation for the end-user. The implementation of Web Services establishes a standard for the service provider to publish data and for the end-users to discover data and services provided; Web Services are solutions to interoperability, data publishing and discovery.

In parallel to setting up the OPeNDAP server, the CO-OPS' IOOS team will need to develop several Web Services and make them available on-demand. Since this will be an ongoing process, whenever there is a need for new functionalities, these functionalities will be offered through Web Services. The CO-OPS' IOOS team will write a set of APIs describing these Web Services and make them available on the IOOS portal. This will give developers and programmers easier access to CO-OPS data programmatically, thus, increasing application interoperability.

To implement Web Services, the CO-OPS' IOOS team recommends Axis, the next-generation SOAP (Simple Object Access Protocol). Axis requires installation of other software and packages such as Xerces, an XML Parser, XML Security API, JavaMail, and JavaBeans Activation Framework.

The CO-OPS' IOOS team will develop a web-based application that will serve as an interface for end-users to browse the available Web Services. Available through this interface will be the set of APIs, along with examples of how end-users can write client-programs to access CO-OPS Web Services. Any updates, as well as suggestions and feedback regarding additional features and services the end-users would like to have will be communicated through this web interface.

Milestone #3: Formalizing Metadata

Formalizing metadata enables CO-OPS to deliver more complete metadata about Currents and Water Level along with other ancillary data. Standardizing its metadata in compliance with the national and international standards will allow CO-OPS to generate, validate and maintain

metadata. CO-OPS will be able to provide a mechanism to ensure that metadata found during discovery are up-to-date, consistent and understandable. In addition, CO-OPS will be able to provide a mechanism to search on metadata, discover the data itself and browse the information of interest.

Syntactic and semantic descriptions of CO-OPS data will be provided, which aid in the discovery, retrieval and use of the data. Metadata will be provided using XML and XML style sheets in accordance with the IOOS recommendations. DMAC has given some guidance as to the standards to follow in creating the metadata, such as the use of the U.S. Federal Geographic Data Committee (FGDC) content standards. CO-OPS should evaluate additional/alternative standards such as Marine XML, Earth Science Markup Language (ESML) and Ecological Metadata Language (EML) to determine which standards best-fit CO-OPS data collection, processing and storage formats.

Some underlying technologies will be needed for the metadata definition process, namely, XML and XML APIs, XSLT and XALAN or Xerces –XML parser. These technologies are needed to be able to manipulate, parse and present XML documents.

The CO-OPS' IOOS team will develop a web-based application to serve as an interface for end-users to browse, search and view the available metadata and their schemas. End-users will have the option to download the XML schemas or view them in a variety of formats. This data search capability will be a powerful tool that is currently lacking on all CO-OPS web pages. This capability will open up an entire level of exposure of CO-OPS data and allow the users to query and select information of interest.

5. System Design

To accomplish the above milestones, new hardware, software and a range of expertise from different areas are needed.

- **Hardware**

A new server is needed to host the OPeNDAP server. The OPeNDAP server will also provide the Web Services to the users and host any necessary web applications.

The CO-OPS' IOOS team has chosen Dell's 1855 Blade System with dual 3.6 GHZ processors, two GB of main memory and dual 300 GB disk space as the platform for this project.

- **Software**

Red Hat ES 3.0 was installed as the operating system on the Dell machine. Most of the software needed for this project will come from the open source community. Using open source allows for more uniformity among the different IOOS servers, promotes sharing of information and keeps costs low. The open source software that CO-OPS will install are Apache as a web server, Tomcat as the Servlet engine, and THREDDS as the OPeNDAP server. Tomcat can be used as a stand-alone server for all the applications to be implemented and developed.

Additional software packages and APIs to develop Web Services and process XML documents and schemas will be installed on the Tomcat server. Axis, the next generation SOAP, and Xerces (an XML Parser), XML Security API, JavaMail, and JavaBeans Activation Framework will also be installed on Tomcat.

- **Overall Architecture**

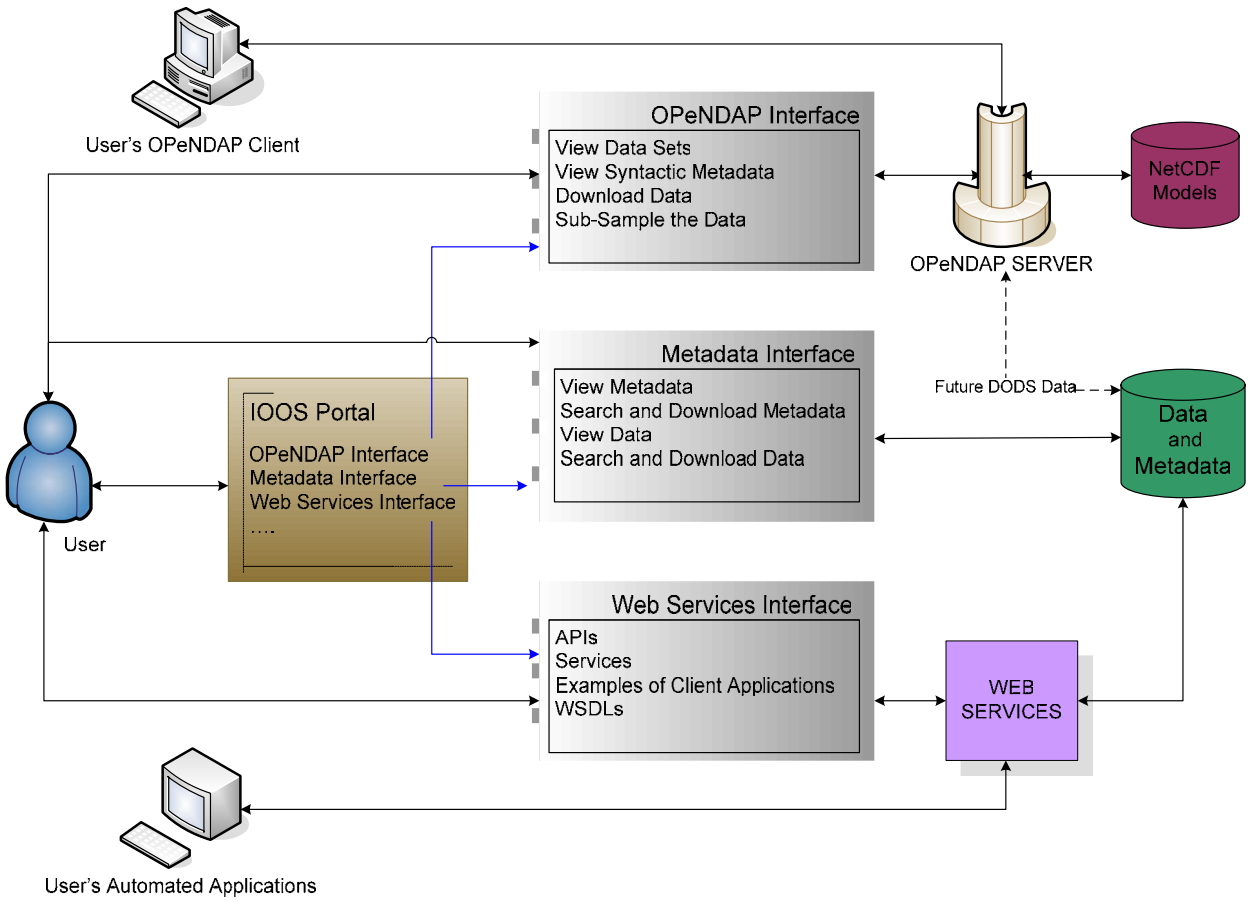


Figure 1: Overall architecture of the IOOS system applications

As shown in figure 1, users can connect either through the IOOS portal or directly to three interfaces of the three applications. Through the OPeNDAP interface, the users of the system can view different datasets, their syntactic metadata, and data. The users can also download the datasets or sub-samples of interest. Through the metadata interface, the users can browse or search the metadata or associated data. Finally, through the Web Services interface, users can discover the APIs and WSDL files and the web services offered.

The OPeNDAP client will allow for automated download of the DODS data, while Web Services will enable users' remote applications automated access to other data.

6. Expertise

This project requires the expertise of Oceanographers, System Administrators and Web Developers.

- **Oceanographers**

Oceanographers are needed to provide the metadata that will go in the information pages as descriptions of the datasets and models. The Oceanographers will also select which models to be served initially, as well as the datasets to be served at a later stage.

Oceanographers' and scientists' input are needed to help gather the requirements for a set of APIs that covers all of the services and products to be provided as Web Services.

- **System Administrators**

System administrators' knowledge about hardware installation, networking and familiarity with the CO-OPS system environment setup is required.

Expertise in system administration is also needed for setup of the environment variables, management of the security of the web server, backup of its data and maintenance requirements, and to identify security protocols to secure the services offered via HTTP (Hyper Text Transport Protocol).

- **Web Developers**

Experienced Web developers who know J2EE technology are needed. A thorough understanding of the Web Services and the XML technologies is ideal since that knowledge may shorten the development time.

The CO-OPS' IOOS team will also need an experienced IT staff member to handle the installation and configuration of THREDDS. This person will also make sure the server can see the appropriate datasets.

7. Testing

The CO-OPS' IOOS team will write a Project Test Plan to define responsibilities and identify test methodologies, phases and environments.

8. Security

IT security is a crucial subject that plays a role in the choice of technology, implementation process and the method of providing services. Security should be managed through access control, roles and quality of the data itself in an effort to preserve the integrity of the system.

Most of the technology recommended and described in this document is based on the same enabling technology that underlies the web, namely HTTP. As a result, all of the common technologies that secure web applications work equally well for this project. Security will consist of firewalls to protect the server and basic authentication. SSL (Secure Sockets Layer) may be used to connect to the server. Other security measures may be taken as needs arise.

9. Assumptions

The described milestones are a first attempt at breaking down the necessary and **preliminary** steps toward implementing the IOOS vision. In achieving these tasks there may be other required sub-tasks that are unknown at this time, especially given the fact that all of these milestones require new technologies and are bound by the ongoing process of setting up standards, definitions and practices of not only the web community but also by the IOOS DMAC steering committee.

This technical plan focuses on the major milestones, and does not go into detail regarding additional components needed to achieve them. Those components could include entire systems, either integrated or self-contained applications. These applications may require gathering requirements, design and implementation plans. The development of such components could require an entire software life cycle that would need to be detailed in separate technical plans. For example, to implement the search capability to query metadata and the data associated with it is an entirely separate application that would be needed for the “Formalizing Metadata” milestone.

10. Challenges

The CO-OPS’ IOOS team expects the milestone of formalizing metadata to pose the most challenges, including the challenge to fully understand, at least at an abstract level, the present state of CO-OPS metadata, and choose a metadata standard for CO-OPS data.

Some technical challenges may be encountered, since most of the technology recommended here will be used for the first time in CO-OPS. These may include the implementation of the recommended technologies, technical support and maintenance.

Regarding expertise, CO-OPS has all the in-house experts needed to accomplish the milestones described in this proposed technical plan. It will be important to make sure that these resources are available and committed to the project vision and to cooperating with the CO-OPS IOOS team.

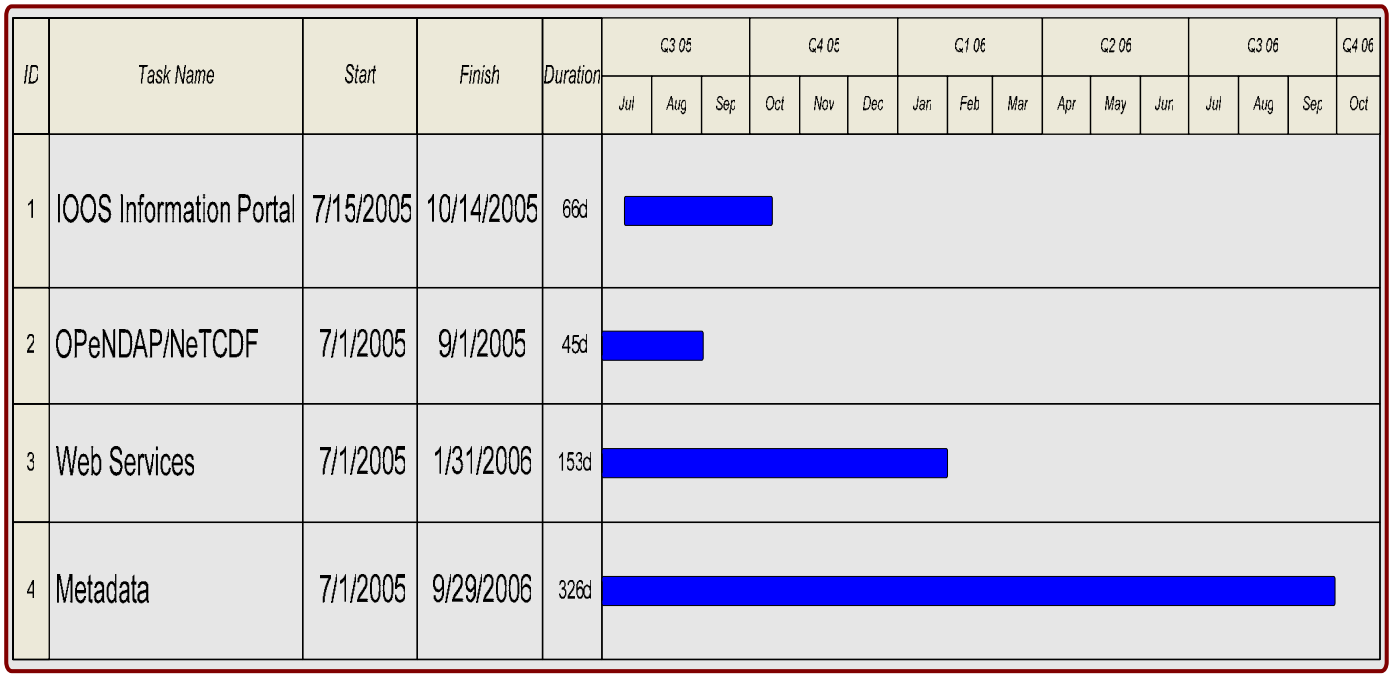
11. Documentation

The CO-OPS’ IOOS team will create internal documentation for all code used during these steps, and external (not associated to the code) documentation in order to provide a historical record of the system through the different phases. These documents will be written in accordance with ISD convention and the Department of Energy’s Systems Engineering Methodology.

The documentation will consist of a combination of web pages and word documents, including as necessary:

- **Installation guides**: documentation that keeps track of the how-to, installation steps, and tips and lessons learned.
- **Requirements** gathered for implementation of a given application.
- **Design and prototypes** needed for web applications development.
- **APIs** that are developed for web services; these serve as valuable documentation tools for internal and external use.
- **Reports**: after the completion of each milestone, the CO-OPS' IOOS team will write a report describing the process, evaluating its phases and status along with detailed information for future reference.
- **Internal documentation**: all the code associated with the backbone of any application will be commented.
- **Project Test Plan**: this document will describe the testing activities of the entire system along with the guidelines for the testing cases to implement.

12. Project Schedule



13. References

National Office for Integrated and Sustained Ocean Observations
http://ocean.us/
The Apache Software Foundation
http://www.apache.org
The OPeNDAP Home Page
http://www.opendap.org
Thematic Real-time Environmental Distributed Data Services
http://my.unidata.ucar.edu/content/projects/THREDDS/index.html
Java Technology Home Page
http://java.sun.com
The World Wide Web Consortium Standards for Web Services
http://www.w3.org/2002/ws/
The Federal Geographic Data Committee Metadata
http://www.fgdc.gov/clearinghouse/clearinghouse.html
Marine Metadata Interoperability site
http://marinemetadata.org/
MarineXML Web Page at The Australian Oceanographic Data Centre implements
http://www.aodc.gov.au/products/prod/marinexml.html
U.S. Department of Energy (2002.) Systems Engineering Methodology Version 3 (DOG G 200.1-1A.) Washington, D.C.: Author.

14. Glossary

API……Application Program Interface
ASCII……American Standard Code for Information Interchange
CO-OPS……Center for Operational Oceanographic Products and Services
DODS……Distributed Oceanographic Data System
DMAC……...Data Management and Communications
FGDC…….Federal Geographic Data Committee
HAZMAT…….…Hazardous Materials Emergency Response
HTML……Hypertext Markup Language
HTTP…….Hyper Text Transport Protocol
IOOS…….…Integrated Ocean Observing System
J2EE……...Java 2 Platform Enterprise Edition
NetCDF…….Network Common Data Format
NOAA ……..National Oceanic and Atmosphere Administration
OPeNDAP…….Open-source Project for a Network Data Access Protocol
SSL……Secure Sockets Layer
THREDDS……Thematic Real-Time Environmental Distributed Data Services
UNIDATA…….University Atmospheric Data access project
WSDL……Web Services Description Language
XML……..eXtensible Markup Language
XSLT…….eXtensible Style sheet Language Transformations